A Research Stream on Sentiment Analysis

TREO Talk Paper

Bo Yuan
Missouri University of Science and Technology
byvf9@mst.edu

Keng Siau
Missouri University of Science and Technology
siauk@mst.edu

Abstract

Sentiment analysis (SA) is a powerful mining technique to study online reviews and comments (Lee & Siau, 2001; Adeborna & Siau, 2014; Zhao & Siau, 2017). It is an advanced text mining technique and the goal of SA is to recognize and extract meaningful information from data using natural language processing (NLP) and computational linguistics.

SA has been applied to areas such as marketing, online social media, customer service, education, and even energy fields (Yuan & Siau, 2017). For example, SA can be used to identify the attitude of customers according to polarity of the reviews and comments that they left online. The data may be from tweets, blogs, or new articles.

Standard SA techniques have been used in many studies and one key component of SA is the lexicon. In SA, building/finding the appropriate lexicon is an important step. Lexicon/corpus selection/construction is generally viewed as a prerequisite for successful SA. Some existing lexicons include the Harvard Inquirer, Linguistic Inquiry and Word Counts, MPQA Subjectivity Lexicon, Bing Liu's Opinion Lexicon, and SentiWordNet.

In this research stream, we study the effect of coupling a specialized lexicon to a general lexicon and its effect on the SA results. The premise is that since lexicon is the most important component in SA, increasing the quantity and strengthening the quality of items in the lexicon will produce better results, especially for specialized domains.

In the first phase of the study, we apply the idea to the petroleum industry and constructed a petroleum domain lexicon for the research. The preliminary results support the hypothesis that coupling a general lexicon with a specialized lexicon enhances the SA.

This stream of research is of values to both academic researchers and practitioners. For academic research, this modular approach provides an efficient and effective way to developing specialized lexicons for various industries and domains. Instead of building an entire lexicon from scratch, only the specialized lexicons need to be built and the specialized lexicons can be coupled to a general lexicon. This "lego pieces" approach is efficient and has been shown to be effective. For practitioners, they can use this approach to quickly build up the lexicons for their industries and specific domains, and they can be confident that combining the specialized lexicons to the general lexicon has been shown to be effective.

References

Adeborna, E., Siau, K. 2014. "An Approach to Sentiment Analysis – The Case of Airline Quality Rating," Pacific Asia Conference on Information Systems (PACIS 2014), Chengdu, China, July 12-14, 2014.

Lee, S., Siau, K. 2001. "A Review of Data Mining Techniques," Industrial Management & Data Systems, (101:1), pp. 41-46.

Yuan, B., Siau, K. 2017. "Lexicons in Sentiment Analytics," *Twelve Annual Midwest Association for Information Systems Conference (MWAIS 2017)*, Springfield, Illinois, May 18-19, 2017.

Zhao, W., Siau, K. 2017. "Machine Learning Approaches to Sentiment Analytics," Twelve Annual Midwest Association for Information Systems Conference (MWAIS 2017), Springfield, Illinois, May 18-19, 2017.